

Analysis of PEAQ Model using Wavelet Decomposition Techniques

Jagadeesh B.K¹, Dr. B Siva Kumar²

¹HKBKCE/Department of ECE, Bangalore, India
Jbkanade_99@yahoo.com

²Dr.AIT/Department of TCE, Bangalore, India
Sivabs2000@yahoo.co.uk

Abstract— Digital broadcasting, internet audio and music database make use of audio compression and coding techniques to reduce high quality audio signal without impairing its perceptual quality. Audio signal compression is the lossy compression technique. It converts original audio signal into compressed bitstream. The compressed audio bitstream is decoded at the decoder to produce a close approximation of the original signal. For the purpose of improving the coding this work attempts to verify the perceptual evaluation of audio quality (PEAQ) model in BS.1387 using wavelet decomposition techniques. Finally the comparison of masking threshold for sub-bands using Wavelet techniques and Fast Fourier transform (FFT) will be done.

Index Terms— Psycho-Acoustic, ATH, DWT, FFT.SMR, CB, WFB, PEAQ

I. INTRODUCTION

Data compression refers to way of reducing data size without affecting the quality of the data; Audio compression is form of data compression. To acquire compressed audio, different audio compression methods have been contrived and implemented. These methods vary from simple technique to most advance and complex that takes sensitivity of the human ear.

In the process of audio compression perceptual limitation of human ear is exploited. This Limitation in human hearing system is used to remove perceptually irrelevant audio signal. MPEG Audio compression technique algorithm achieves compression by exploiting the perceptual limitation of the human ear. By applying audio compression algorithms it is possible to get compact digital representations of audio signals for efficient transmission without impairing the quality at the receiving end. The main purpose of the audio compression is to represent the audio signal with a less number of bits while achieving transparent signal reproduction.

The absolute threshold of hearing (ATH) is used to characterize the amount of energy needed in a pure tone such that it can be detected by a listener in a noiseless environment. The absolute threshold is typically expressed in terms of dB SPL. The frequency dependence of this threshold was quantified as early as 1940, when Fletcher reported test results for a range of listeners that were generated in a National Institutes of Health study of typical American hearing acuity. The quiet (absolute) threshold is well approximated by the nonlinear function [1][6].

Absolute threshold of hearing is used to shape the coding distortion spectrum is the first step toward perceptual coding. Absolute threshold is of limited value in the coding context. Finding threshold for spectrally complex quantization noise is a modified version of the absolute threshold, with its shape

DOI: 01.IJRTET.10.2.1365

© Association of Computer Electronics and Electrical Engineers, 2014

determined by the stimuli present at any given time. Since stimuli are in general time-varying, the detection threshold is also a time-varying function of the input signal. Auditory masking is a psychoacoustical phenomenon in which a weak signal is masked in the presence of a stronger signal, the stronger signal is called masker and the signal which is masked by stronger signal is called maskee. Exploiting this phenomenon in perceptual audio compression is achieved so that the original audio signal is treated as a masker for distortions introduced by lossy data. A masking threshold is computed based on the frequency representation of the signal.

More specifically, the Fast Fourier Transform (FFT) coefficients are used to evaluate the masking threshold. The psychoacoustic model used in the perceptual audio coder is based on the Psychoacoustic Models. The MPEG-1 Audio Standard describes two different psychoacoustic models, i.e. psychoacoustic models-1 and psychoacoustic models-2 the first being computationally simpler and suitable for coding at higher bit rates and the second being more complex but also more reliable at lower bit rates.

An auditory model was developed by the International Telecommunications Union (ITU) within the framework of the Perceptual Evaluation of Audio Quality (adopted as ITUR BS.1387) [9][18]. PEAQ provides advanced metrics for the assessment of the perceptual quality of audio signals. Among other model output variables, a masking threshold is estimated from the auditory model.

The majority of MPEG [3] coders applies a psycho-acoustic model for coding and uses the filter bank to approximate the frequency selectivity of the human auditory system. Figure1 (a) and Figure1 (b) shows a diagram of the structure of a generic perceptual audio coder. Figure 1(a) shows the structure of the encoder, which has three main stages and a fourth is bit stream formatting stage and Figure 1(b) shows the decoder, which has three stages.

The encoded audio signal (compressed audio) acts as an input to the decoder and decoder reconstructs the original signal from the encoded bitstream. The three stages in the decoder, is the reverse operations of encoder. Three stages in the encoder. Namely, the signal analysis, Quantization and encoding, and bitstream formatting stages of the encoder correspond to the signal synthesis, de-quantization and decoding, and bitstream extraction stages of the decoder, respectively. The additional block in the encoder is the psychoacoustic model, which is not required in the decoder since the information is encoded as side-information. This means perceptual coders are asymmetrical and the encoder has a more complex and requires more computations than the decoder. In this work the author verifies the masking threshold (energy) of subbands for PEAQ model using wavelet decomposition techniques instead of FFT (conventional method)

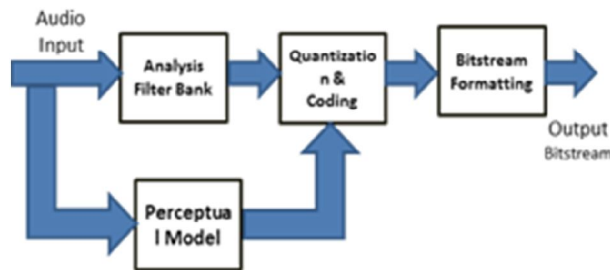


Figure1 (a) Encoder



Figure1 (b) Decoder

The discrete wavelet transform can conveniently decompose the signal into an auditory critical band-like partition [2][11]. Signal decomposition into critical bands resulting from Wavelet analysis needs to satisfy the spectral resolution requirements of the human auditory system

A. Wavelet Decomposition

Wavelet decomposition [4][13] provides a solution that makes possible for a finer an adjustable resolution of frequencies at high frequencies. This makes adaptation to particular signals [5]. Psychoacoustic model achieves an improved decomposition of the signal into critical bands(CB) using the discrete wavelet decomposition transform (DWT). This results in a spectral partition which approximates the critical band distribution much closer than before. Furthermore, the

Masking thresholds are computed entirely in the Wavelet domain to get approximation of critical bands. Wavelet analysis should meet the spectral resolution requirements. The wavelet basis also plays important role to satisfy temporal resolution of the signal. The continuous wavelet transform (CWT) of signal x relative to the basic wavelet is given by:

$$W_{\psi} x(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (1)$$

Where a, b ($a, b \in \mathbb{R}$, $a \neq 0$) are respectively the translation and scale parameters. Furthermore, $\psi(t-b/a)$ represents the wavelet basis functions that are derived from a single mother wavelet function, $\psi(t)$, through dilations a and translations b . The wavelet basis functions represent an Orthonormal basis to the space of $L^2(\mathbb{R})$ such that,

$$L^2(\mathbb{R}) = \text{span} \{ \psi_{ab}(t); a \in \mathbb{R}^+, b \in \mathbb{R} \} \quad (2)$$

If the basic wavelet satisfies the admissibility condition, then the wavelet reconstruction formula is:

$$x(t) = \iint_{\mathbb{R}} W_{\psi} x(a, b) \psi_{a,b}(t) \frac{da db}{a^2}$$

The Wavelet filter bank (WFB) is a filter bank that offers a great deal of flexibility in terms of the choice of the basis filter and the decomposition tree structure. Additionally, the WFB offers a variety of ways of handling boundary artefacts in the context of block processing. The following sections describe the design of the WFB in terms of these broad design “parameters”.

The standard DWT involves a dyadic tree structure in which the low-channel side is successively split down to a certain depth. Wavelet tree decomposition is a wavelet transform in which the signal is passed through more number of filters. The detail coefficients will be obtained from the right-leaf node of each level and the approximation coefficients will be obtained from the left-leaf node at the lowest level. Fig.2 illustrates DWT where the nodes represent the wavelet coefficients (at various decomposition levels).

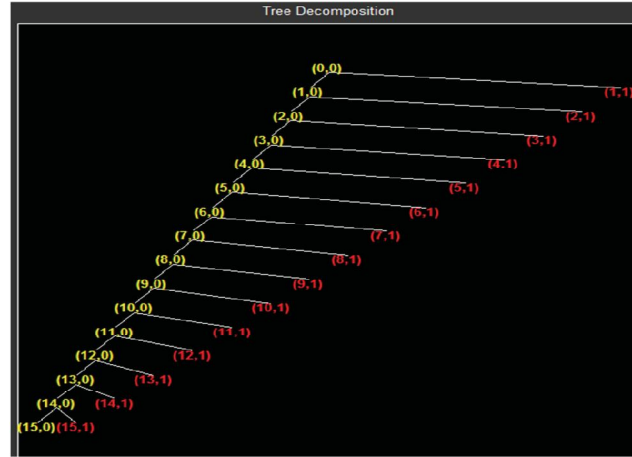


Figure2 .Wavelet Tree Decomposition

Wavelet tree decomposition with depth one splits the signal into high pass and low pass bands. With depth two will splits the low pass spectrum from depth one. Each stage wavelet tree decomposition splits low pass spectrum from previous stage, this yields an octave band pass filter bank wherein sampling rate of each subband is proportional to its bandwidth. Wavelet analysis is efficient because of portions of the frequency towards the low frequency the psychoacoustic model [7] is based on many studies of human perception. Studies have proven that the average human does not able to hear all frequencies as same.

Figure 3 shows the wavelet tree decomposition. Here “C” denotes the coefficients in the various decomposed branches of the tree and “L” denotes the number of the coefficients in the corresponding nodes of tree. Another type of decomposition is wavelet packet decomposition.

Figure 4 shows the wavelet packet decomposition with depth three. Wavelet packet decomposition with depth one splits the signal into high pass and low pass bands and With depth two will splits the low pass spectrum from depth one. Each stage wavelet packet decomposition splits low pass from previous stage in to low pass and high pass spectrum and each high pass from previous stage in to low pass and high pass. This yields an octave band pass filter bank wherein sampling rate of each subband is proportional to its bandwidth.

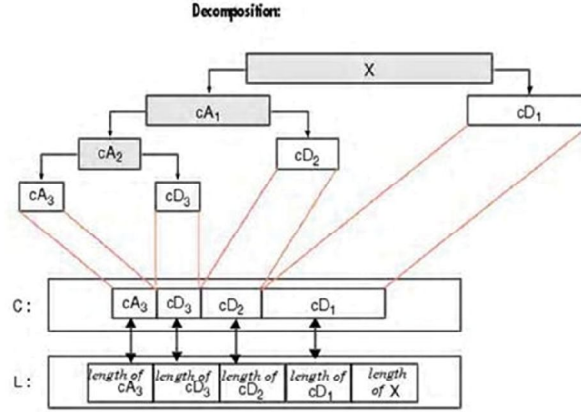


Figure 3: Wavelet tree decomposition

Wavelet analysis is efficient because of portions of the frequency towards the low frequency the psychoacoustic model is based on many studies of human perception. Studies have proven that the average human does not able to hear all frequencies as same. While choosing specific wavelet decomposition the author have considered some restrictions to create orthogonal translates and dilates of the wavelet (the same number of coefficients than the scaling functions), and to ensure regularity (fast decay of coefficients controlled by choosing wavelets with large number of vanishing moments). In this work author used orthogonal [15][16] family of wavelets with name “daubechies (DB10)”.

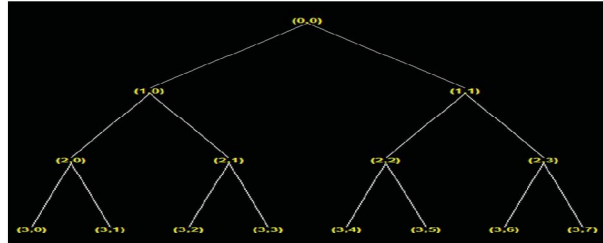


Figure 4: Wavelet Packet decomposition

B. PEAQ model

PEAQ (Perceptual Evaluation of Audio Quality) is the objective measurements Recommendation Standard of perceived audio quality established by ITU in 1998, which is also called BS.1387. It utilizes software to simulate perceptual properties of human ear, and then integrates multiindices to evaluate subjective quality of test audio PEAQ measurement method models fundamental properties of the auditory system [17]. Several intermediate stages of the standard models physiological and psychoacoustic effects. In this paper author has done the changes in the BS.1387 with respect to the human ear model without changing the functionality of the standard. In this paper the work is related human ear model using the wavelet decomposition techniques instead of FFT and other parts of the standard is unchanged. Figure 5 illustrates the block diagram PEAQ model, This work also aims at comparing the making energies obtained from the proposed method and the

FFT, as mentioned in the standard. BS.1387 involves both ear model and cognitive model. This work aims at the ear model only and cognitive model is not the scope of this work.

In MPEG psychoacoustic model 2, SMR (ratio between signal energy and masking threshold) is determined by experiential value of examination observation. The Basic version of PEAQ adopted a design approach different from MPEG audio standard. It attempts to combine physiological structure of human ear with masking effect of simple signal represented from examination to find the inherent consequence, and then use mathematic model to emulate the structure of human ear. Figure 5 illustrates the block diagram of this design [10], in which the function of outer/middle ear, inner ear, and audio perception related nerve cell and brain are emulated. This kind of psychoacoustic model could be extended conveniently to acoustic masking of complex signal.

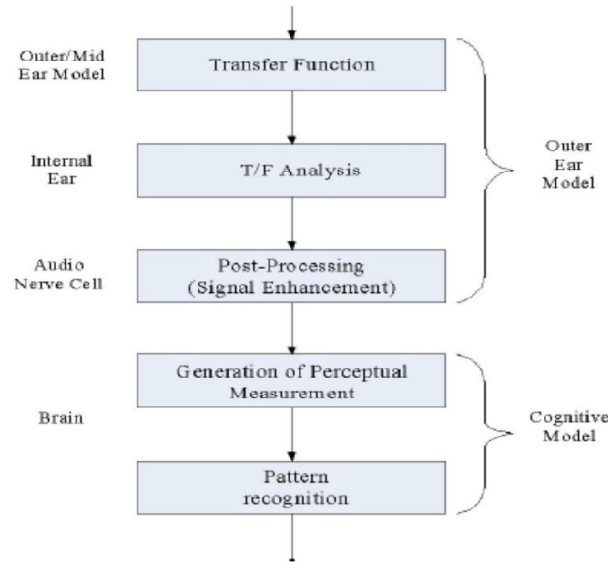


Figure 5: psychoacoustic model based on PEAQ (Basic version)

PEAQ was initially designed to measure audio quality without special consideration for requirements of audio coder, such as window switching, unification of critical band scale, estimation of masking properties on transient signal, etc. Thus, we must take into account these requirements and correct the psychoacoustic model of PEAQ basic version so as that it could be used in audio coder. Figure 6 shows the proposed design of basic version of PEAQ psychoacoustic model [14].

The process of human perception is modelled by employing a Difference measurement technique that compares reference [12] Signal and a test signal (i.e. the "output" signal of the codec)

II. PROPOSED DESIGN USING WAVELET DECOMPOSITION TECHNIQUES

Method 1:

Using Wavelet Tree Decomposition

The step involved in the proposed method is as shown in figure 6. The implementation steps as below.

Step 1: Framing:

The ".wav" file is actually an uncompressed audio signal, the sampling rate chosen here is 48000Hz. Divide the audio signal into different frames each frame of size 2048

Step 2: Apply Wavelet tree decomposition to each Frame and Calculate the Energy:

Apply wavelet packet of depth level seven will give total of 128 subbands that replicate the 109 critical bands in PEAQ model(Basic Version) consider only 109 subbands for the analysis. The wavelets used in this work are Debusis wavelets.

Step 3: Outer and Middle Ear Modelling

The outer and middle ear response is modeled by the frequency dependent weighting function[17]

Step 4: Frequency Grouping

In this step take a frame of frequency domain samples and group them into the frequency bands these frequency bands will be grouped according to the critical bands

Step 5: Adding of internal Noise

Internal noise is generated by blood flow within the human ear and this will be modelled by adding the frequency dependent Energies of the frequency groups. The internal noise function is implemented with the reference[17].

Step 6: Frequency Spreading

The spreading function is adopted from an auditory model developed by Terhardt [18]

Step 7: Time domain spreading

Temporal masking effect is incorporated in time domain Spreading.

In this model temporal masking is considered by means of first-order smooth filtering [17].

Step 8: Calculation of Masking Parameters

The masking threshold is calculated by weighting the excitation patterns [17].

The masking (Threshold) energy for each band is obtained by

$$E_{mask(dB)}(i) = E_{f(dB)}(i) - m_{dB}(i)$$

Method2:

Using Wavelet Packet Decomposition

The step involved in the proposed method is as shown in figure 7. The implementation steps as below

Step 1: Framing:

The “.wav” file is actually an uncompressed audio signal, the sampling rate chosen here is 48000Hz. Divide the audio signal into different frames each frame of size 2048

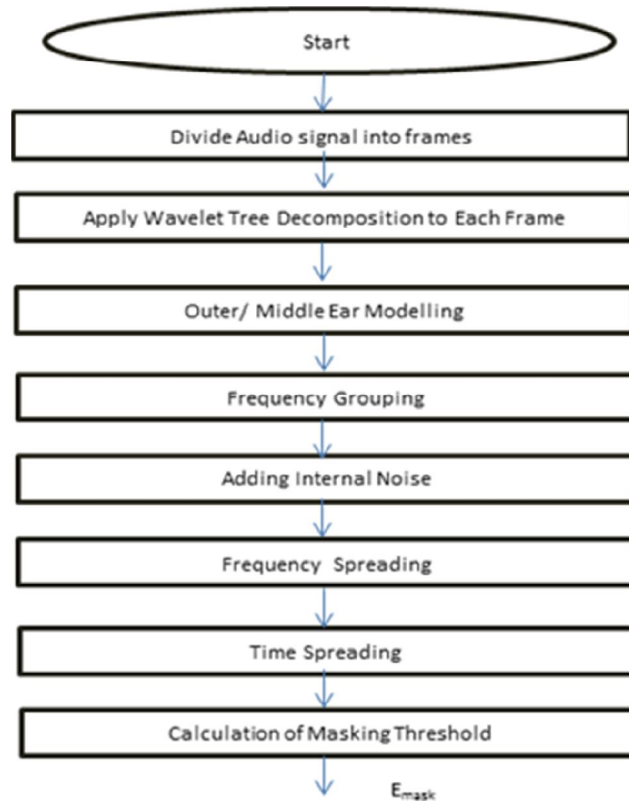


Figure 6: Flow chart for the proposed PEAQ model (Method 1)

Step 2: Apply Wavelet packet decomposition to each Frame and Calculation of the masking Energy:

Apply wavelet packet decomposition for each frame and the depth level seven will give total of 128 subbands that replicate the 109 critical bands in PEAQ model (Basic Version). The wavets used in this work is Debusis wavelets

Step 3: Follow the step 3 to step8 from method 1.

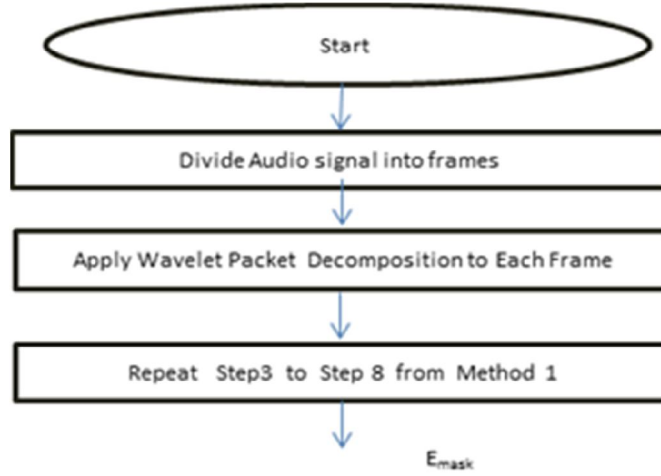


Figure 7: Flow chart for the proposed PEAQ Model (Method 2)

A. Experimental Results

In this work we have implemented PEAQ model using Wavelet decomposition techniques. The two decomposition techniques used in this paper are wavelet tree decomposition (Method 1) and wavelet packet decomposition (Method 2).

Author has used several inputs to test the proposed algorithm. The results of the two inputs have listed in this paper. Figure 8, figure 9 and figure.10 indicate the results of test vector one and figure 11, figure12 and figure 13 indicate the results for test input 2. For the test input 1 Figure 8 shows the implementation using FFT. Figure 9 and Figure 10 shows the results of proposed wavelet techniques. From the Figure 9 and Figure 10 it is evident that masking energy obtained from the proposed methods better estimate the masking energy for subbands under consideration compare to FFT method. Figure 9 and figure.10 gives the better masking threshold estimate compare to that of masking threshold obtained from the FFT method (Figure 8). For the test input 2 same observations can be made from the figures 11, 12 and figure 13.

Test input 1: Results

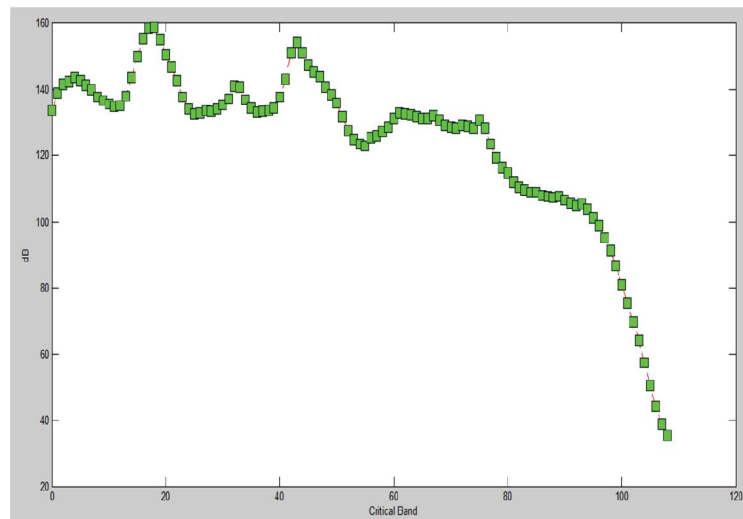


Figure 8 : Distribution curve of masking energy using

FFT (PEAQ Basic Version)

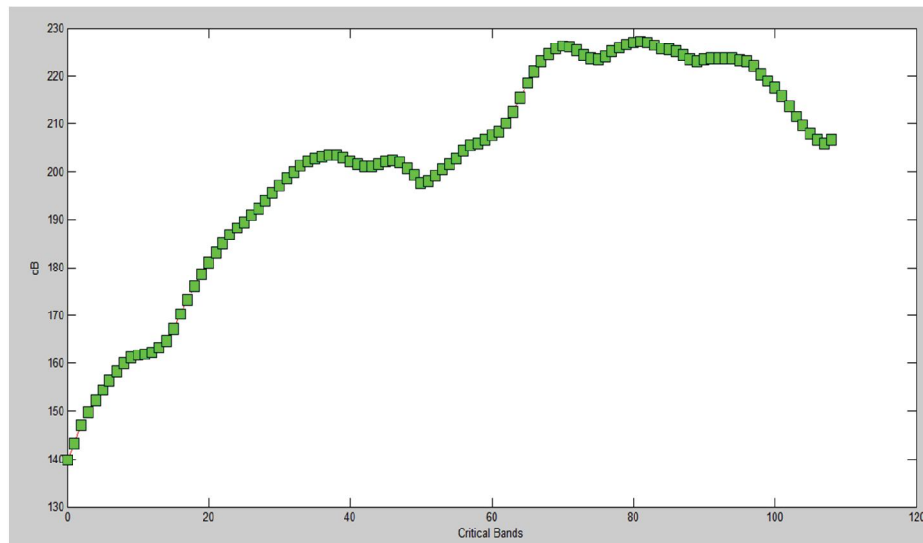


Figure 9: Distribution curve of masking energy using Proposed method (Wavelet packet Decomposition)

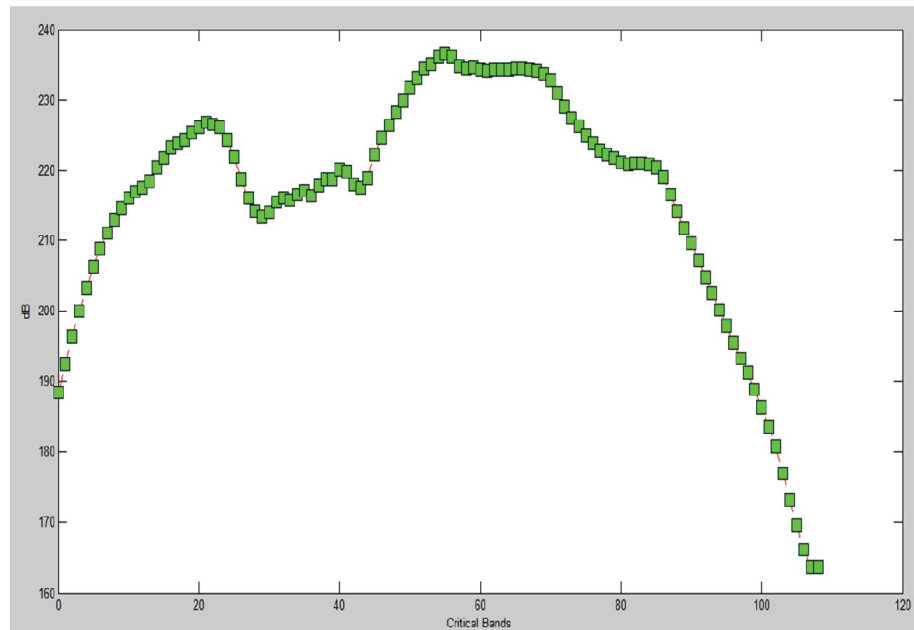


Figure 10: Distribution curve of masking energy using Proposed method (Wavelet Tree Decomposition)

Using wavelet decomposition techniques we can observe that only for first few sub bands the the masking threshold is more compare to FFT methods but as we are moving towards higher number of subbands the masking energy is predominantly less comapre to that of FFT .Hence it will contribute towards the data compression.Again among the defferent wavelet decomposition techniques Wavelet tree decomposition gives highre values of masking compare to that of wavelet packet decompostion method for the same subabnds under consideration. From the results we can observe that wavelet tree decomposition technique is giving more masking energy compare to FFT and wavelet packet decomposition techniques.

Test input 2:Results

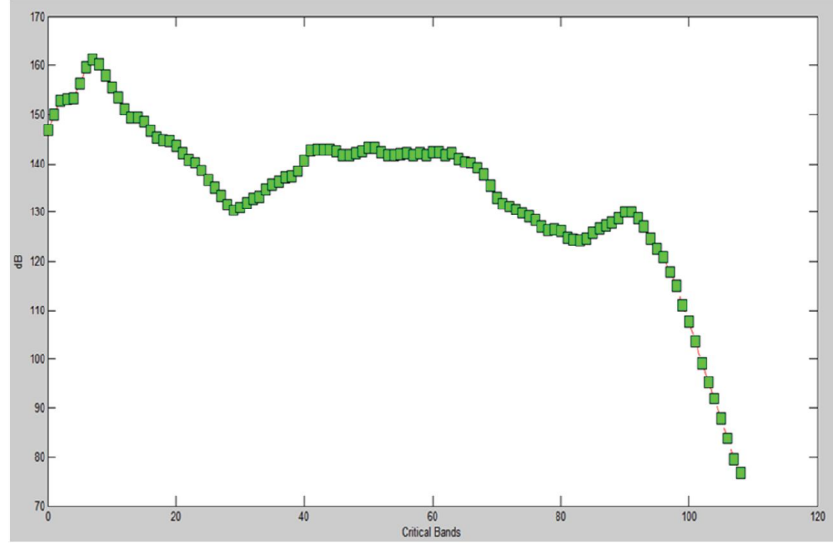


Figure 11: Distribution curve of masking energy using FFT (PEAQ Basic Version)

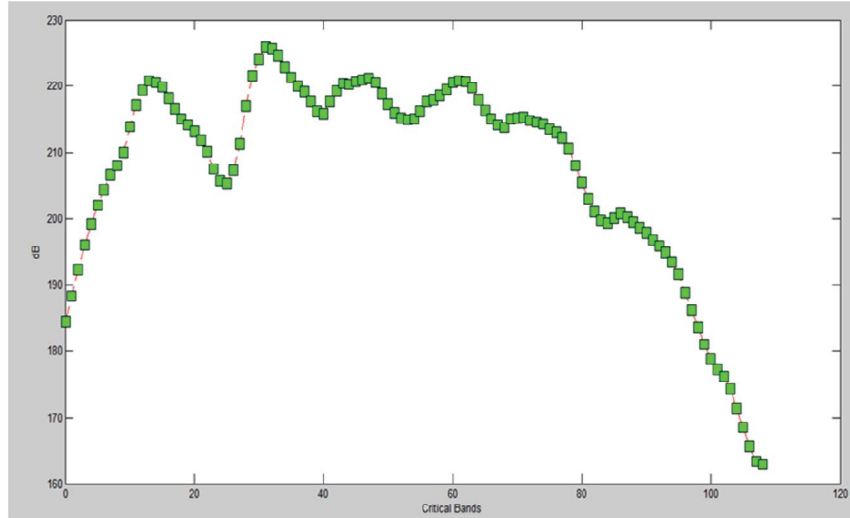


Figure 12: Distribution curve of masking energy using Proposed method (Wavelet Tree Decomposition)

III. CONCLUSION AND FUTURE WORK

The PEAQ based psychoacoustic model using wavelet decomposition techniques takes an account of the critical bands and masking phenomenon. The specialty of the proposed techniques is that it gives an analysis by wavelet decomposition on the frequency bands that gives the closer approximation of the critical bands of the ear. In estimates masking threshold more accurately compare to masking threshold obtained from FFT as in the standard BS.1387 standard.

The future work involves the analysis of PEAQ model using the wavelet lifting scheme and audio quality test (As mentioned in the other parts of the standard BS.1387) which involves the integration of the proposed method with the cognitive model of the standard BS.1387. Also the proposed method can be integrated with the other blocks of the MPEG audio codecs to get overall compression ratio.

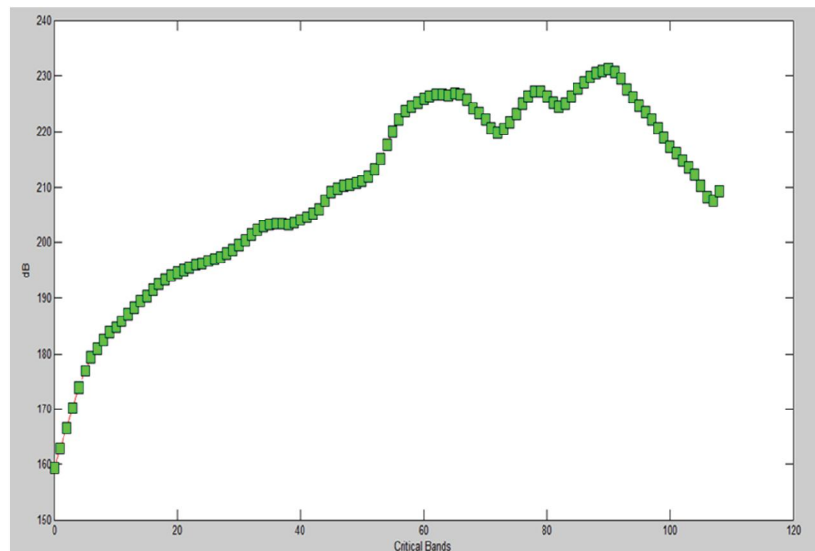


Figure 13: Distribution curve of masking energy using Proposed method (Wavelet packet Decomposition)

REFERENCES

- [1] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE Trans.Sig.*, vol. 88, Apr. 2000, pp. 51-513
- [2] D.Sinha and A.Tewfik, "Low Bit Rate Transparent Audio Compression using adapted wavelets," *IEEE Trans.Sig.Proc.*, pp.3463-3479, Dec 1993.
- [3] Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1, 5 Mbit/s-Part3: Audio, 1999.
- [4] M. V. Wickerhauser, "Adapted wavelet analysis from theory to software," Wellesley, Massachusetts, 1994.
- [5] C.S.Burrus, R.A. Gopinath and H.Guo, "Introduction to wavelets and wavelets transforms: A Primer," Prentice Hall, 1998.
- [6] H.Najafzadeh-Azghandi, "Perceptual Coding of Narrowband Audio Signals." PhD thesis, McGill University, Montreal, Canada, April 2000.
- [7] M. R. Zurera, F. L. Ferreras, M. P. J. Amores, S.M. Bascón, and N. R. Reyes, "A new algorithm for translating psycho-acoustic information to the wavelet domain," *Signal Processing*, vol. 81, no. 3, pp. 519-531, 2001.
- [8] B. Lincoln, "An experimental high fidelity perceptual audio coder," Project in MUS420 Win97, March 1998.
- [9] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 314-323, 1988.
- [10] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PEAQ - the ITU standard for objective measurement of perceived audio quality," *J. Audio Eng. Soc.*, vol. 48, pp. 2-29, Jan. 2000.
- [11] B. Carnero and A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms," *IEEE Transactions on Signal Processing*, vol. 47, no. 6, pp. 1622-1635, 1999.
- [12] Dermot Campbell, Edward Jones, Martin Lavin "Audio quality assessment techniques—A review and recent developments" *Elsevier Signal Processing* 89 (2009) 1489-1500
- [13] Strang, G., Nguyen, T., 1996. "Wavelets and Filter Banks" Wellesley-Cambridge Press, Wellesley, Massachusetts.
- [14] Xiaopeng Hu Guiming He Xiaoping Zhou Peaq-Based Psychoacoustic Model for Perceptual Audio Coder, ICAOT2006
- [15] Meyer, Y. (1989), "Orthonormal Wavelets", in *wavelets*, Proc. Int. Conf. France, Springer, Berlin.
- [16] Ogden RT. (1997), "Essential Wavelets for Statistical Applications and Data Analysis". Birkhauser, Boston.
- [17] Kabal, P. 2002. An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality. Montreal, Department of Electrical & Computer Engineering, McGill University, TSP Lab technical report. 89 p.
- [18] International Telecommunication Union, Method for Objective Measurements of Perceived Audio Quality, July 1999. ITU-R Recommendation BS.1387.